CLAIMS

That which is claimed:

1.   A method, comprising:

identifying an event having an associated article;

identifying article data associated with the article; and

determining a capture score for the event based at least in part on the article data.

2.   The method of claim 1, further comprising determining a threshold value.

3.   The method of claim 2, further comprising compiling event data associated with the event if the capture score is above the threshold value.

4.   The method of claim 1, wherein a capture score is determined at least in part by associating a weight with one or more fields of an event schema.

5.   The method of claim 1, wherein the article data comprises a location of the article.

6.   The method of claim 5, wherein the capture score is determined at least in part by associating a weight with the location of the article.

31

7.   The method of claim 6, wherein the weight is determined at least in part by user behavior.

8.   The method of claim 1, wherein the article data comprises a file type of the article.

9.   The method of claim 8, wherein the capture score is determined at least in part by associating a weight with the file type of the article.

10.   The method of claim 6, wherein the weight is determined at least in part by user behavior.

11.   The method of claim 1, wherein the article data comprises access data associated with the article.

12.   The method of claim 11, wherein access data comprises recency associated with access of the article.

13.   The method of claim 11, wherein access data comprises frequency associated with access of the article.

14.   The method of claim 1, wherein the capture score is determined at least in part by associating at least one weight with the article data.

15.   The method of claim 14, wherein the weight is determined at least in part by user behavior.

16.   The method of claim 2, wherein the threshold value is predetermined.

17.   The method of claim 2, wherein the threshold value is determined based at least in part on user behavior.

18.   The method of claim 2, further comprising indexing the event if the capture score is above the threshold value.

19.   The method of claim 2, further comprising storing the event if the capture score is above the threshold value.

20.   The method of claim 2, further comprising not indexing the event if the capture score is below the threshold value.

21.   The method of claim 2, further comprising not storing the event if the capture score is below the threshold value.

22.   The method of claim 1, wherein the event is a historical event.

23.   The method of claim 5, wherein the location of the article can comprise a directory identifier in which the article is stored.

24.   The method of claim 1, wherein the article is identified during a crawl of a client device.

25.   The method of claim 1, further comprising determining if the article meets at least one criterion and not capturing the event if the article meets the criterion.

26.   A computer-readable medium containing program code, comprising:

program code for identifying an event having an associated article;

program code for identifying article data associated with the article; and

program code for determining a capture score for the event based at least in part on the article data.

27.   The computer-readable medium of claim 26, further comprising determining a threshold value.

34

28.   The computer-readable medium of claim 27, further comprising program code for compiling event data associated with the event if the capture score is above the threshold value.

29.   The computer-readable medium of claim 26, wherein a capture score is determined at least in part by associating a weight with one or more fields of an event schema.

30.   The computer-readable medium of claim 26, wherein the article data comprises a location of the article.

31.   The computer-readable medium of claim 30, wherein the capture score is determined at least in part by associating a weight with the location of the article.

32.   The computer-readable medium of claim 31, wherein the weight is determined at least in part by user behavior.

33.   The computer-readable medium of claim 26, wherein the article data comprises a file type of the article.

34.   The computer-readable medium of claim 33, wherein the capture score is determined at least in part by associating a weight with the file type of the article.

35.   The computer-readable medium of claim 34, wherein the weight is determined at least in part by user behavior.

36.   The computer-readable medium of claim 26, wherein the article data comprises access data associated with the article.

37.   The computer-readable medium of claim 36, wherein access data comprises recency associated with access of the article.

38.   The computer-readable medium of claim 36, wherein access data comprises frequency associated with access of the article.

39.   The computer-readable medium of claim 37, wherein the capture score is determined at least in part by associating at least one weight with the article data.

40.   The computer-readable medium of claim 39, wherein the weight is determined at least in part by user behavior.

41.   The computer-readable medium of claim 27, wherein the threshold value is predetermined.

42.   The computer-readable medium of claim 27, wherein the threshold value is determined based at least in part on user behavior.

43.   The computer-readable medium of claim 27, further comprising program code for indexing the event if the capture score is above the threshold value.

44.   The computer-readable medium of claim 27, further comprising program code for storing the event if the capture score is above the threshold value.

45.   The computer-readable medium of claim 27, further comprising program code for not indexing the event if the capture score is below the threshold value.

46.   The computer-readable medium of claim 27, further comprising program code for not storing the event if the capture score is below the threshold value.

47.   The computer-readable medium of claim 26, wherein the event is a historical event.

48.   The computer-readable medium of claim 30, wherein the location of the article can comprise a directory identifier in which the article is stored.

49.   The computer-readable medium of claim 26, wherein the article is identified during a crawl of a client device.

50.   The computer-readable medium of claim 26, further comprising program code for determining if the article meets at least one criterion and not capturing the event if the article meets the criterion.

51.   A method, comprising:

    identifying an event having an associated article by crawling a client device;

    identifying one or more of a location of the article, a file type of the article, and access data for the article;

    determining a capture score for the event based at least in part on one or more of the location of the article, the file type of the article, and the access data for the article; and

    indexing the event if the capture score is above a threshold value.

52.   A system, comprising:

    a means for determining a capture score for an event; and

    a means for indexing the event if the capture score is above a threshold value.

53.   The system of claim 52, further comprising a means for storing the event if the capture score is above the threshold value.